Cigil – A Framework for Image Transferring over Sound encoding

Mehmet Kutlu
Electronics and Communication Department
Istanbul Technical University
Istanbul, Turkey
kutlumeh@itu.edu.tr

Ahmet Haluk Açarçiçek
Electronics and Communication Department
Istanbul Technical University
Istanbul, Turkey
acarcicek@itu.edu.tr

***Abstract*—We present Cigil, a framework for transferring image data over sound encoding. And framework offers data transfer without any connection. Cigil is meant to be used by non-expert users and compatible with everyday use.**

I. Introduction

Transfering and sharing images became everyday use lately. People transfer image almost every second of the time. But this transfers requires any type of connection between devices, e.g. internet, bloutooth. Therefore, the idea of transfering images with no connection is very important. In this paper, we present Cigil, our framework that transfers image over sound.

By making this transfer possible over sound mapping, no connection will be needed. With this development, image transfering will be possible without any cost.

II. Protocol

To transfer an information via sound a solid protocol is necessary and the rules ,which contained by protocol, must be solid and robust to maket he transfer more error proof , easy to implement and open to new improvments.So communication protocol is the key of our communication system Cigil.

We implement the system to images so our protocols block diagram contains important parts of the Cigil.

Figure 1. Block diagram of system

A. Image

Avarage high quality images today can be up to 8 megapixel but to transfer those kind of huged sized pictures without loosing information needs high bandwidth transfer methods but in our protocol bandwith is limited to hearable sound (0-20000Hz). So in Cigil first we reduce size of image to 128x128 px. This is an option but according to theoretical limits huge sizes will be transfered huge amount of time.

Figure 2. Sample Rgb image

Next step to reduce size of data is removing the colors of image. Because of this is not a commercial product we keep things simple as possible. This optimized formula is used to convert the image to grayscale considering human visual system.

GrayscaleImage=0.2989*RedChannel+0.5870*GreenChannel+0.140*Blue Channel

Equation 1. RGB to gray scale

Final step in the image processing is the quantizing step. In cigil 256 gray level is quantized to 8 gray level without equalizer.

B. Image to sound mapping

In Cigil we represent 8 quantized gray levels to different sound frequencies and append a sinus wave sound at mapped frequency at desirable pixel frame time. So to represent 128x128 grayscale image as sound minimum time can be found according to equation 2.

$$\text{Sound Time} = \text{Total pixel}(128 \times 128) \times \text{Pixel Frame}$$

Equation 2. Total sound time

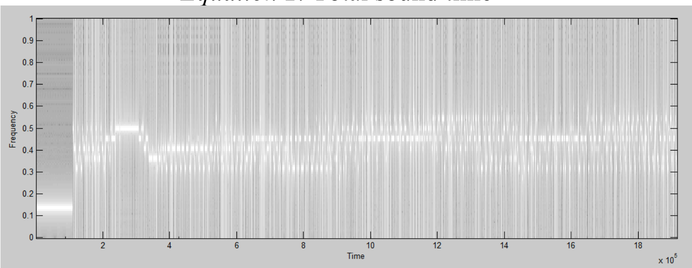


Figure 3. Created sound signal

Also pixel frame can not be lower than theoretical communication limits which we will mention at theoretical limits section.

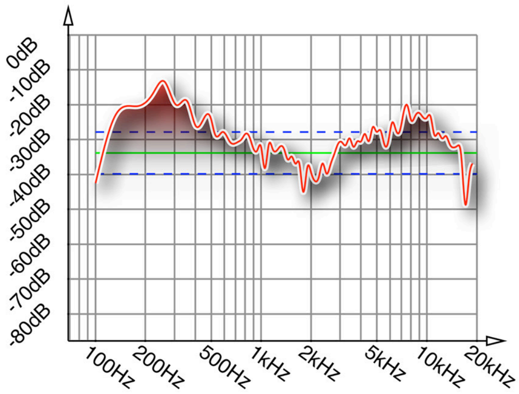


Figure 4. Avarage microphone frequency response

To map quantized pixel levels to frequency couple of things must be considered. Firstly avarage mobile smart phone and laptop microphone frequency response can be seen in Figure 4.

Those microphones designed to obtain frequency level similar to human audio spectrum. Generally people talks between voice frequency (300- 3000Hz) so to make Cigil more robust against noisy places we didnt use that voice band and according to the avarage microphones frequency response ideally we use between 6000Hz to 13000 Hz.

Gray Level	Mapped Frequency
1	6000Hz
2	7000Hz
3	8000Hz
4	9000Hz
5	10000Hz
6	11000Hz
7	12000Hz
8	13000Hz

C. Decoding

Firstly decoder listens to the environment and when sycronizing signal is heared decoder understands that after the syncronizing signal, which is 3000 Hz default in Cigil , decoder anlysis the data using STFT(Shot Time Fourier Transform) and finds peak frequencies in giving pixel frame. Basicly at according to sample rate and given pixel frame decoder algorithm calcutes the amount of fourier window per pixel and assign median of the peak frequencies to that pixel.

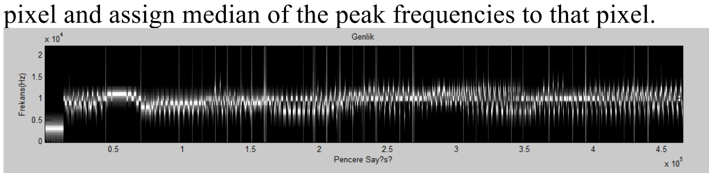
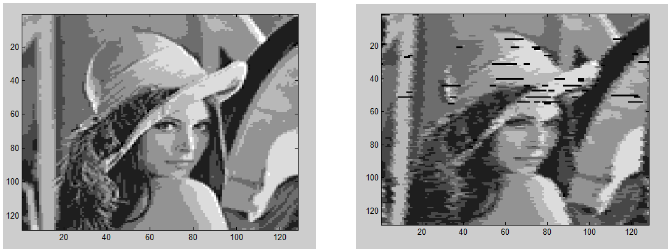


Figure 5. Reconstructed spectrogram

After that Cigil, maps frequencies back to grayscale levels and reconstruct images.



(a) (b)

Figure 6. (a) Reconstructed image (b) Reconstructed image

D. Synchronization

The most significant problem is the synchronization between the transmitter and receiver. To solve this problem one constant frequency is chosen and send for 5 seconds. On the receiver side, it listens the sound and searches the frequency in the protocol to start getting pixel values. 100 sample length frame is used for searching the frequency on the received sound. After finding the last frame includes known frequencies 5 sample length frames searches the last frequency sent by the transmitter and starts to get pixel values.

III. Theoretical Limits

Default sampling rates of everyday use devices are 44.1kHz. Therefore, maximum available frequency is 22kHz. On this project, 8-layer quantization is used and 1kHz interval choosed to be durable to frequency interference. And when decoding sound back to image with sampling rate 44.1kHz, to keep frequency resolution lower but closer to 1kHz, 128 window size for hanning window function and %50 overlap is used to calculate STFT(short time fourier transform).

$$w(n) = 0.5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right)$$

Equation 3. Hanning window function

$$STFT = X(\tau, w) = \int_{-\infty}^{\infty} x(t) \times w(t-\tau) e^{-jwt} dt$$

Equation 4. Short time fourier transform

That size is close to 1ms, for this reason minimum duration for sound of each pixel must be greater that or equal to 1ms. And to be resistant to error 5ms duration is used.

IV. CONCLUTION

REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] Douglass L. Mansur, Meera M. Blattner and Kenneth I. Joy, ``Sound Graphs, A Numerical Data Analysis Method for the Blind,'' *Journal of Medical Systems*, Vol. 9, pp. 163-174, 1985.

[5] P. B. L. Meijer "An experimental system for auditory image representations", *IEEE Trans. Biomed. Eng.*, 1992

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.